Sentiment Analysis of Opinions over Time Toward Saudi Women's Sports

Norah J. Almateg(✉), Sarah M. BinQasim, Jawaher N. Alshahrani, Ahad Y. Marghalani, and Zahyah H. Alharbi

Management Information Systems Department, College of Business Administration, King Saud University, Riyadh, Saudi Arabia
{442920287,442920306,442920303,442920320}@student.ksu.edu.sa, zalharbi@ksu.edu.sa

Abstract. The examining and promoting women's health is essential to the general well-being of society. As sports activities are important for maintaining physical, psychological, and social health, female participation in sports is one of the significant factors in achieving the goals of Saudi Arabia's Vision 2030. In this paper, sentiment analysis was used to compare the society opinions pre and post permit-ting Saudi women's sports, between 2017 and 2021. To identify the sentiment of a given tweet, a lexicon for the Saudi dialect was developed. In total, 12,000 tweets were collected and prepared. After data preparation, the tweets were reduced to 1,999 across all selected hashtags for this the initial study. We used four different hashtags related to Saudi women's sports, namely, (#Officially_female_sports_in_schools) represented as a Pre-Hashtag, whereas (#Women_Sport) and (#Female_Sport) as Pre and Post Hashtags, and (#Tahani _Alqahtani) as Post-Hashtag. The data in each hashtag were classified as positive, negative, or neutral. To build the sentiment classifier model, A Support Vector Machine (SVM) classifier was applied. The highest average accuracy was for the Pre-Hashtag with a score of 91%, followed by the Pre and Post Hashtag with a score of 85%. Finally, the Post Hashtag has the lowest score of 72%. The results show that 81% of the sample are positive. Accordingly, women have been becoming more motivated to engage in sports participation, as well as Saudi society is being more encouraging.

Keywords: Women's sport · Sentiment analysis · Natural Language Processing (NLP)

1 Introduction

Women's sports history started back in the nineteenth century, particularly when the second Olympic games took place in 1900 and female athletes participated for the first time [1]. Women's sport has always been recognized in some countries, but in countries such as Saudi Arabia, women's sport has been a controversial subject for many years. In

S.-H. Wang and Y.-D. Zhang (Eds.): ICMTEL 2022, LNICST 446, pp. 247–259, 2022.
https://doi.org/10.1007/978-3-031-18123-8_19

recent years, Saudi Arabia has been witnessing the rise of female athletes, and women are increasingly becoming open about participating in sports compared to previous years.

To analyze opinions in society toward women's sport, sentiment analysis was used in this research. Sentiment analysis involves using Natural Language Processing (NLP) techniques, computational linguistics, and text analytics to identify and extract subjective information from the source materials, aiming to determine the attitude of a speaker or a writer toward a certain topic or incident [2].

The Saudi community has seen an increase in the use of social media platforms such as Twitter [3]. Twitter is a powerful tool for disseminating information and an excellent source of opinionated text about a wide range of topics: politics, buisness, economic, and social. As a result, the NLP research community has become interested in studying this rich language resource. To the best of the author's knowledge, no previous study has used sentiment analysis to analyze Twitter users' opinions about women's sports in Saudi Arabia. Therefore, this initial study seeks to compare the society opinions pre and post permitting Saudi women's sports.

This paper is organized as follows: Sect. 2 covers the history of women sports in Saudi Arabia; Sect. 3 discusses related studies; Sect. 4 describes the study's methodology; Sect. 4.3 addresses the sentiment analysis process; Sect. 5 presents the results of the analysis; and Sect. 6 includes the conclusion and the future work.

2 History of Saudi Women's Sports

In 2013, Saudi Arabia's first dedicated sports center for females was opened in Al-Khobar city [4]. That same year, females in private schools were officially allowed to engage in sports, while in public schools, it was not permitted until 2017 [5]. The decision to allow physical education in 2017 was made by Saudi Arabia's Ministry of Education (MoE) in order to fulfill the goals of the Saudi Vision 2030 and promote healthy practices in society. Women in 2018 were also allowed to attend events at sports stadiums [6]. As part of Saudi Arabia's push to advocate for more female participation in sports and develop a more inclusive sports environment, the country in 2017 also appointed Princess Reema Bint Bandar, the first women in the kingdom to take on such an important role as the head of the Saudi Federation for Community Sports (SFCS) [7]. In her role as Vice President for Development and Planning at the Saudi Arabian General Sports Authority, the Princess has supported female participation in sports as well as contributing to the development of an inclusive sports environment for women. She has also worked on legitimizing women's gyms and focused heavily on encouraging women to use the streets and public parks to exercise. As she stated, "I've been telling women they don't need permission to exercise in public; they don't need permission to activate their own sports programs. And more and more, they are doing it. The choices that women have today are greater than yesterday and every day they will grow more" [8].

3 Related Work

In this section, previous research articles are summarized to give readers a general overview of women's sports in Saudi Arabia. Following this, studies that have used sentiment analysis in Arabic are examined.

3.1 Women's Sports in Saudi Arabia

Alruwaili in [9] studied the relationships between sport, gender, education, region, and religion in order to provide recommendations on how to improve gender inclusivity in sport in Saudi Arabia. The study focused on answering four questions. First, what are the dominant attitudes towards women's participation in sport in Saudi Arabia? Second, what are the key social, cultural, and civic issues that affect women's participation in sport in Saudi Arabia? Third, how do different interpretations of Islam influence attitudes towards women's sport in Saudi Arabia? Fourth, how are the ideas about women's participation in sport in Saudi Arabia changing? The researcher used a mixed-methods approach. In particular, an exploratory survey was undertaken via a questionnaire that was distributed to 890 individuals (444 responded: 196 women and 248 men). In addition, thousands of tweets from the micro-blogging site Twitter were examined with a representative sample of 96 selected for discussion. A thematic analysis of both the interview and Twitter data was performed. Alruwaili concluded that there is support for women to participate in sport and physical activity, most frequently on health grounds. This support was not explicitly constrained by male authority or Islamic teachings, although religiously-motivated reasoning was apparent in a proportion of the sample. More specifically, support was relatively high in the data across education levels, city of origin, and gender for women to participate in sport and physical activity. However, one restriction on this participation that most respondents across all data-gathering methods agreed upon was that women's participation in sport should be in accordance with the teachings of Islam, sex-segregated, and occur in a private setting.

Another study on women's sports was undertaken by Fakehy, Alfadhi, and Alotaibi [10]. The study focused on identifying the factors affecting the attitude of undergraduate female students in Saudi Arabia toward sports. The relationship between physical fitness, social experience, formal competition, physiological experience, and sports attitude of female students was examined. The researchers used questionnaire data to validate their hypotheses, which were as follows: H1) Physical fitness has a relationship with psychological experience; H2) Physical fitness has a relationship with sports attitude. 645 female students responded to the survey. After analyzing the data, the authors found that physical fitness, social experience, and formal competition positively influenced physiological experience, which in turn had a positive effect on the attitudes of female students toward sports.

Al-Haramlah, Merza, and Albakerin [11] investigated the level of physical activity among Saudi women and explored differences in terms of factors such as place of residence, age, weight, educational level, profession, and marital status. This study used an interventional approach to support efforts that encourage physical activity in Saudi women. A pilot sample of 80 females was interviewed to secure the validity and reliability of the preliminary instrument. The study was driven by two main questions. First, what

are the attitudes of Saudi women toward practicing physical activities? Second, are there any significant statistical differences in Saudi women's attitudes towards practicing physical activities that are based on the provinces of residence, age group, weight, educational level, nature of profession, or marital status? The researchers concluded that Saudi women in general are interested in practicing physical activities if opportunities and suitable facilities are available; they positively support the idea of providing women's sports clubs, as well as the idea of incorporating physical activities within girl schools. The findings also revealed that there are statistical differences in Saudi women's attitudes toward engaging in physical activity that are attributable to their provinces of residence, age, weight, educational level, and the nature of the profession.

Sayyd, Zainuddin, Zulaika, and Altowerqi in [12] discussed the current state of physical activity and sports activities in the Kingdom of Saudi Arabia (KSA) compared to other countries in the world. The authors also examined the issues that hinder the success and development of physical and sports activities development in KSA. The researchers conducted their review of the literature using ScienceDirect, Springer, the Journal of Health Sciences, and Google Scholar databases. A total of 58 articles were included in their research. The study revealed that the main factors hindering the success of physical and sports activities in the KSA's universities among males were lack of energy, motivation, self-confidence, and time, while among females, the factors were lack of social support and resources. The researchers concluded that there was no motivating environment that encouraged sports participation. They also found that there is limited action and initiative in terms of sports activities and sports participation in KSA compared to many other countries worldwide.

The research undertaken by Al-shahrani [13] aimed to determine the extent of sports practices among women in Saudi society by identifying the associated motives and obstacles to participation. Moreover, the researcher sought to develop suggestions and solutions to advance the culture of sports practice among women in Saudi society. The study focused on answering three questions. First, what are the motives for practicing sports for Saudi women in society? Second, what are the obstacles to women's sports in Saudi society? Third, what are the proposals and planning indicators for activating sports for women in Saudi society? The researcher conducted a questionnaire with 432 Saudi women in Riyadh. The study was based on a descriptive approach. Al-shahrani concluded that Saudi women are generally eager to maintain their fitness and increase their motivation for sports practice. The high cost of using fitness centers was one of the obstacles identified for women's sport participation in society. In addition, the absence of school curricula for sport reduces Saudi women's awareness of the importance of practicing sport.

Notably, several methods were used in the studies mentioned in this literature review, but none has applied sentiment analysis. Therefore, we found it worthwhile to apply sentiment analysis to analyze Arabic tweets in order to study Saudi society opinions pre and post permitting women's sports.

3.2 Arabic Sentiment Analysis

Al-Twairesh, Al-Khalifa, Al-Salman and Al-Ohali in [3] aimed to define a methodology that could be used in collecting and constructing a large dataset of Arabic tweets. Their

research was motivated by the lack of sufficient resources that allow the application of Arabic sentiment analysis. The extracted dataset contained almost 2.2 million tweets and was used to generate an Arabic corpus of tweets. The researchers cleaned and preprocessed the collected dataset by first filtering retweets, URLs, and mentions, and then by removing Arabic letters using normalization and tokenization. After manual annotation, the corpus was reduced to 17,573 tweets. Four labels were used: positive, negative, neutral, and mixed. The corpus included tweets written in Modern Standard Arabic (MSA) and the Saudi dialect. Three annotators were recruited to resolve the conflicts in annotation through majority voting. A list of guidelines for annotation was defined. After completing the annotation, a questionnaire was developed to evaluate the methodology. The results showed that two annotators stated that the guidelines were clear. The annotators were asked if the annotation of tweets was clear; all three annotators chose sometimes. The annotators were also asked which label was the hardest to determine; two annotators said mixed and one annotator said neutral.

Another study on Arabic sentiment analysis undertaken by Alqmase, Al Muhtaseb, and Raabaan [14] aimed to build a classification model using sentiment analysis by formulating sports Arabic text into fanatic and anti-fanatic contexts. Anti-fanatic text was defined as text that helps to decrease sports-fanaticism, while fanatic text increased sports-fanaticism. This was achieved by formalizing the social text into 21 fanatic and anti-fanatic contexts using proposed indicators. Fanatic indicators were aggression, agitation, hatred, and passion, whereas anti-fanatic indicators were adaptation, knowledge, respect, affections, and intimacy. Then, the authors developed a fanatic lexicon with 1,780 terms. After that, 919,000 domain-specific tweets were collected and labeled. To build a classification model, machine learning algorithms were applied. As a result, the best-built classifier achieved 91% accuracy. The proposed classification model can help governments to measure the impact of their efforts to reduce sports-fanaticism.

Ali in [15] aimed to conduct a comprehensive emotion mining and sentiment analysis task during the pandemic by collecting Arabic tweets related to online learning. The author extracted data using Twitter APIs, where the collected tweets were associated with seven different hashtags related to COVID-19. Then, the data were prepared for intensive preprocessing. This included removing hashtags, URLs, identifying emoticons, user mentions, and extra spaces. Also, punctuation was replaced with a single space. Spelling correction was applied to prepare the dataset for stemming. In addition, specific letters were normalized and stop words were removed. All emojis and emoticons provided by Twitter were kept and considered as a part of the texts. The most frequent emojis were defined, after which every emotion was replaced with its typical weight using the NRC lexicon. The different emotion annotations for a target term were consolidated by selecting the emoji with the highest weight. Two different datasets were used for the experiment. The datasets were collected between 20 September 2020 and 15 October 2020, and the total number of records across both datasets was 10,487. Finally, different classification algorithms were applied, including Naïve Bayes (NB), Multinomial NB (MNB), K-Nearest Neighbor (KNN), Logistic Regression (LR), and Support Vector Machine (SVM). The results showed that the proposed model performed well in analyzing people's perceptions about the coronavirus, achieving an accuracy of 89.6% using SVM classification. As for emotion analysis, anger was found to dominate the

tweets, followed by the fears surrounding the first attempt to engage in distance learning. This was mostly due to the lack of face-to-face communication, network system failure, ambiguity.

4 Methodology

This section presents the research methodology, which consisted of three phases: data collection, data preparation, and sentiment analysis, as shown in Fig. 1.

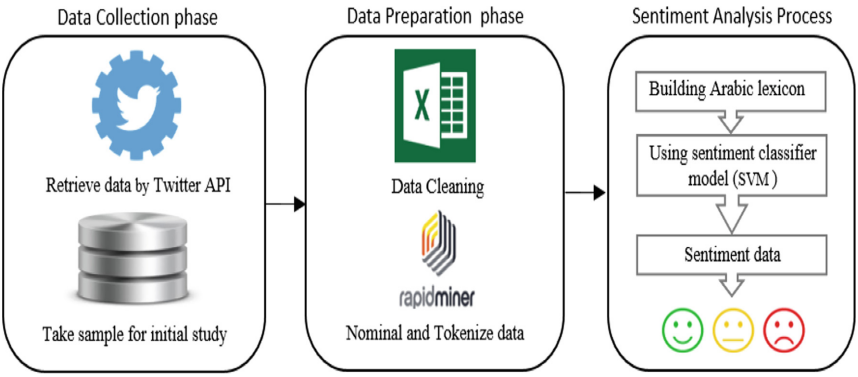

Fig. 1. The applied Arabic sentiment analysis methodology.

4.1 Data Collection

Data were retrieved from Twitter using the Python Tweepy library and an API key. 12,000 Arabic tweets were extracted using four different hashtags. For each hashtag,

Table 1. Hashtags used to form dataset

Dataset ID	Hashtag	Type	Date	Size
S01	#رسميا_رياضه_البنات_بالمدارس Official-ly_female_sports_in_schools	Pre	July 2017	670
S02	#رياضة_البنات #Female_Sport #الرياضة_النسائية #Women_Sport	Pre/ Post	2017 – 2021	660
S03	#تهاني_القحطاني #Tahani _Alqahtani	Post	30 July 2021	669

tweets were collected over a different period in order to compare the society opinions pre and post permitting Saudi women's sports, as well as changes in sentiment over time. The dataset size and period for each hashtag were identified as shown in Table 1.

4.2 Data Preparation

Since the text in tweets is known to be noisy, it should be cleaned and pre-processed before being analyzed. We started by removing the noisy tweets that contained ads, coupons, political materials, and unrelated tweets. We also excluded tweets that were posted on a date outside the covered period specified in this study. To accomplish the preparation and preprocessing tasks, we used a variety of tools, which are detailed in this section.

Microsoft Excel
We first cleaned the collected data using Microsoft Excel formulas. This involved excluding retweets and tweets containing URLs or media, as well as removing user mentions (@user) from the tweets along with hashtags (#), emojis, and punctuation. Then, we normalized several Arabic letters that have different shapes, including (,ة), by converting them into a single shape. For example, the different forms of "alif" (,,) were converted into ( ), and the letter "ta'a"(ة) was converted to (ه).

RapidMiner
The second step involved using the RapidMiner tool. RapidMiner is a software platform developed by the company that integrates machine learning, data and text mining, predictive analytics, and business analytics into a single platform. It is a sophisticated offering with over 1,500 drag-and-drop operators, allowing for the most efficient and quick data mining operations [16].

For our work, we utilized the "Nominal to Text" operator to convert all nominal attributes to string attributes, after which we used text processing techniques. "Tokenization" was applied, which breaks a document's text into a sequence of tokens. "Filter stopwords" was also used to filter Arabic stopwords from a document by removing every token which matched a stop word in the built-in stop word list. The stemming process was excluded due to its poor outcomes, which works to reduce the inflectional forms of each word into a common base word or root word or stem word. The authors in [17] demonstrated that using the stemming operator produces unsatisfactory results in the Arabic language.

After data preparation, the quantity of tweets across all four hashtags was reduced from 12,000 to 7,500. Then, a limited sample was randomly constructed for the initial study, consisting of 1,999 tweets covering all hashtags. It is worth referring to [18], which proved that the performance of a classification algorithm does not depend on the size of the datasets used.

4.3 Sentiment Analysis Process

Sentiment analysis is a technique of recognizing and extracting subjective information from source materials by Using NLP, computational linguistics, and text analytics. It

involves calculating people's perceptions, attitudes, and emotions regarding entities, persons, situations, events, themes, and their characteristics [19]. This study used a lexicon-based approach to extract sentiments from tweets. An Arabic lexicon was required to label the data before applying the classifier algorithm.

Lexicon-Based Classification

Lexican-based classification is a method of assigning labels to documents based on the number of words in two contrary lexicons, such as negative and positive sentiment [20]. Lexicon-based approaches require the manual collection of opinion words, and so they have been criticized for requiring too much human effort. Furthermore, the Arabic language is dialectically rich and its diverse structural properties in the various dialects need to be fully captured in order to derive maximum benefit from Arabic sentiment analysis, particulary for less formal channels such as social media [21].

It was challenging to find a lexicon that included Saudi dialects related to sports for this study. Two prior studies have constructed Arabic lexicons with the Saudi dialect [22, 23]. Accordingly, a lexicon was built for this study. First, we collected data from Twitter, particularly from the sport context, focusing on the Saudi dialect. Then, we tokenized these data. In turn, we classified the most important words into positive and negative. In the classification process, all authors participated, and then their results were compared to assure the polarity of the classification. Lastly, duplicated words were removed.

As a result of this process, 4,765 were the total number of words in the dictionary. 2,701 words were positive and 2,062 were negative. The dictionary is available on GitHub [24] for other researchers to use in future work.

Experimental Setup

The classification process was performed using the Python programming language (specifically, Jupyter notebook in Anaconda navigator). First, the built lexicon was applied on the dataset to assign the polarity of each word in the tweets to either positive or negative. Each tweet was then labeled based on its maximum polarity by counting the number of positive and negative words in the tweet. For example, if the tweet contained three positive and one negative word, it would be classified as a positive tweet (and vice versa); if the number of the negative words is greater than the positive words, the tweet would be classified as positive. Moreover, if the number of positive and negative words is equal, the tweet is classified as neutral. An illustration is shown in Table 2.

Once the dataset is labeled, Support Vector Machine (SVM) was applied to build the sentiment classifier model. SVM is a collection of supervised learning algorithms used for classification, regression, and outlier detection [25]. We used SVM due to its effective performance in Arabic sentiment analysis, according to [26].

We applied the train-test split() function to split that dataset into 70% training data (to train the classifier) and 30% test data. Moreover, to ensure accuracy and prevent overfitting, 10-fold cross-validation was performed on the SVM. This process was undertaken on all three datasets mentioned in Table 2.

Table 2. Sample of tweets labeling

Dataset ID	Hashtag	Pos count	Neg count	Label
S01	رياضه البنات دس السم العسل Translated to: Female' sport put poison in honey	2	2	Neutral
S02	العقل السليم الجسد السليم Translated to: A healthy mind a healthy body	2	0	Positive
S03	خسرت نفسها قبل ان تخسر المباراة Translated to: She lost herself before losing the game.	1	2	Negative

5 Results and Discussion

To evaluate the performance of SVM methods for each dataset, we used different evaluation metrics: Accuracy = (TP + TN)/(TP + FP + TN + FN), Recall = TP/(TP + FN), and Precision = TP/(TP + FP).

In these three equations, TP (true positive) is the number of correct positive predictions, TN (true negative) is the number of correct negative predictions, FP (false positive) is the number of incorrect positive predictions, and FN (false negative) is the number of incorrect negative predictions [27].

Table 3 shows the results for the SVM model for each dataset. As indicated, SVM demonstrated superior performance in 3-class classification, since it is able to differentiate between sentiment polarity.

Table 3. Results of SVM model

Classifier	Dataset ID	Accuracy	Recall	Precision
SVM	S01	0.91	0.90	0.90
	S02	0.85	0.85	0.85
	S03	0.72	0.71	0.71

In the next subsection, further results are discussed for each dataset, where each dataset represents a certain hashtag during a specific period, as shown in Table 1.

5.1 Pre-hashtag (S01)

To evaluate and confirm the sentiment analysis for society toward Saudi women's sports at the beginning of the decision of permitting sports in female schools. 670 tweets were randomly selected from our dataset as a balanced sample from the hashtag (#رسميا_رياضه_البنات_بالمدارس, which is translated as (#Officially_female_sports_in_schools). This was a trending hashtag on Twitter in July 2017

when the Ministry of Education (MoE) in Saudi Arabia announced that sports classes would soon be introduced officially in female schools.

Table 4 shows that 90% of the sample was satisfied with the introduction of sport and physical activity in female schools, with 604 positive tweets. By contrast, only 10% of the tweets showed negative or neutral sentiments.

Table 4. Comparison of positive, negative, and neutral tweets for pre-hashtag

Dataset ID	Size	Positive	Negative	Neutral
S01	670	604	36	30

According to the results above, it is confirmed that there was a need for female sports classes in schools, referring to [9], which proved that there is support for women to participate in sport and physical activity. However, most of the tweets agreed that women's participation in sport should be undertaken in accordance with the teachings of Islam, sex-segregated, and occur in private settings. On the other hand, the authors in [13] proved that women are interested in maintaining their fitness and increasing their motivation for sports practice. The main obstacle that women face was the high cost of participating in the gym, with the absence of school curricula for sport having been overcome by the decision of the MoE.

5.2 Pre and Post Hashtags (S02)

In this hashtag, 660 tweets were randomly selected from the dataset as a balanced sample in order to evaluate and confirm the sentiment analysis for the pre and post hashtag (#رياضة البنات - #الرياضة النسائية), which translates to (#Female_Sport #Women_Sport). To compare societal opinions over time, two separate periods were selected (2017 and 2021). The results show that 85% of the sample was classified as positive, where some were supporters of women participating in sports, while others were female athletes already participating in sport.

Table 5. Comparison of positive, negative, and neutral tweets for pre and post hashtags

Dataset ID	Size	Positive	Negative	Neutral
S02	660	575	53	32

According to the results in Table 5, women in general are interested in sports participation in the case of the availability of facilities, a suitable environment, and necessary supplies and materials.

As mentioned in [12], that there is no motive environment that encourages sports participation, Therefore, it is noteworthy that in this study, the results suggest otherwise; in particular, comparing 2017 with 2021, women are becoming more motivated to engage

in sport, and the environment is increasingly becoming encouraging. Referring to [10], the study proved that physical fitness, social experience, and formal competition had a positive effect on physiological experience, which in turn positively influenced the attitudes of female students towards sports.

5.3 Post-hashtag (S03)

To evaluate and confirm the sentiment analysis for the post hashtag (#تهاني_القحطاني), which translates to (#Tahani _Alqahtani), 669 tweets randomly were selected from the dataset as a balanced sample.

72% of the results were supporters of Tahani's participation in the Olympics, while the rest criticized her. The results in Table 6 show unrealistic responses arising from the high level of rejection of Tahani's participation owing to religious and political factors, which results in getting out of context in a way that is not related to sport.

Table 6. Comparison of positive, negative, and neutral tweets post-hashtags

Dataset ID	Size	Positive	Negative	Neutral
S03	669	442	118	109

Another study conducted in 2020 [12], which mentioned that females lack social support and resources, is proved by the results above. Tahani Al-Qahtani, a Saudi Women who participated in the Tokyo Olympics in 2021, received different points of view on her participation: some were supporters, while others attacked her due to religious and political affiliations and motivations.

Additionally, some tweets mentioned that Saudi women's activities must be improved. For example, one tweet stated: "There is very little action in sports activities in the KSA compared to other countries in the world." The supportive tweets showed that women are inspired to participate if they have the opportunity and a suitable place, which was mentioned in [11].

6 Conclusion and Future Work

The aim of this study was to compare the society opinions pre and post permitting Saudi women's sports. We collected a dataset of 12,000 tweets that were prepared, consisting of text related to Saudi women's sports from Twitter. The tweets reduced to 7,500 across all four hashtags after data preparation. For the initial study, a limited sample was randomly constructed consisting of 1,999 tweets covering all hashtags. For the analysis process, we performed sentiment analysis using the SVM classifier, which showed a good performance. The highest accuracy was for the Pre-Hashtag with a score of 0.91, followed by the Pre and Post Hashtag with a score of 0.85. Finally, the Post Hashtag has the lowest score of 0.72.

Our results, which compared people's opinions toward women's sports from previous years with nowadays (2017 vs. 2021), in addition to the progress between these years.

The pre opinions dataset showed that people previously wanted women to engage in sport, as well as participate in physical activity during education; therefore, most of the opinions agreed with this decision, suggesting that this would be helpful and healthy for society. In contrast, as of 2021, citizens and residents across Saudi society have already engaged in many kinds of sport. They are also in agreement, support, and are proud of this engagement. This trend is reflected in the result of the third dataset, which is related to a Saudi woman, Tahani Alqahtani, who participated in the Olympics 2021. However, the outcome of the second dataset, concerning pre and post opinions sentiment in 2017 vs. 2021, showed that the societal interest has moved toward greater women's engagement in sport. Moreover, the available data suggest that sport is becoming a major aspect of Saudi women's lifestyles.

This paper is an initial study that analyzes opinions in Saudi society toward women's sports. Future work will continue to consider more hashtags that are recent, as well as comparing them with older dates, to find better results. Another recommendation is to use more general hashtags that target the society more. Also, include thematic analysis to be able to provide recommendations that improve women's sports in Saudi Arabia.

References

1. Vadhera, N.: Historical sketch of women's participation in sports: an overview. Int. J. Yogic Hum. Mov. Sports Sci. **3**(2), 417–422 (2018)
2. Tejwani, R.: Sentiment analysis: a survey. arXiv, pp. 1–3 (2014)
3. Al-Twairesh, N., Al-Khalifa, H., Al-Salman, A., Al-Ohali, Y.: AraSenTi-Tweet: a corpus for Arabic sentiment analysis of Saudi tweets. In: 3rd International Conference on Arabic Computational Linguistics, ACLing, Dubai, UAE, pp. 63–72 (2017)
4. Saudi Arabia opens first sports centre for women. https://gulfnews.com/world/gulf/saudi/saudi-arabia-opens-first-sports-centre-for-women-1.1192220. Accessed 21 Dec 2021
5. Saudi Arabia approves physical education program in girls' schools. https://www.arabnews.com/node/1127811/saudi-arabia. Accessed 27 Dec 2021
6. Saudi Arabia to let women into sports stadiums. https://edition.cnn.com/2017/10/29/middleeast/saudi-arabia-women-sports-arenas/index.html. Accessed 27 Dec 2021
7. Determined and Inspiring: 50 Saudi Women in Sports. https://www.abouther.com/node/15196/people/leading-ladies/determined-and-inspiring-50-saudi-women-sports. Accessed 27 Dec 2021
8. Ladies Leading Team, Princess Reema on the Future of Sports and Women in Saudi Arabia, About Her. https://www.abouther.com/node/9071/people/leading-ladies/princess-reema-future-sports-and-women-saudi-arabia. Accessed 16 Oct 2021
9. Alruwaili, M.: Females and sport in Saudi Arabia: An analysis of the relationship between sport, region, education, gender, and religion. Ph.D., eTheses, Faculty of Health Sciences and Sport, University of Stirling, Scotland, UK (2020)
10. Fakehy, M., AlFadhil, A., Hassan, Y., AlOtaibi, H.: Empowering Saudi women for sports as a response to KSA Vision 2030: attitudes of undergraduate female students toward practicing sports. J. Sports Psychol. **30**(2), 53–66 (2021)
11. AlHaramlah, A., Merza, H., Albakr, F.: Saudi women's attitudes toward physical activity: an interventional approach to improving future health. Asian Soc. Sci. **12**(7), 95 (2016)
12. Sayyd, S., Zainuddin, Z., Diyana, Z., Altowerqi, Z.: Sports Activities for undergraduate students in Saudi Arabia universities: a systematic literature review. Int. J. Hum. Mov. Sports Sci. **8**(1), 1–16 (2020)

13. Al-Shahrani, H.: The reality of women's sport in Saudi society. Int. J. Hum. Mov. Sports Sci. **8**(6), 525–535 (2020)
14. Alqmase, M., Al-Muhtaseb, H., Rabaan, H.: Sports-fanaticism formalism for sentiment analysis in Arabic text. Soc. Netw. Anal. Min. **11**(1), 1–24 (2021). https://doi.org/10.1007/s13278-021-00757-9
15. Ali, M.: Arabic sentiment analysis about online learning to mitigate covid-19. J. Intell. Syst. **30**(1), 524–540 (2021)
16. RapidMiner Documentation Operators. https://docs.rapidminer.com/latest/studio/operators/. Accessed 15 Nov 2021
17. Alami, N., Meknassi, M., Ouatik, S. A., Ennahnahi, N.: Impact of stemming on Arabic text summarization. In: 2016 4th IEEE International Colloquium on Information Science and Technology (CiSt), Tangier, Morocco, pp. 338–343 (2016)
18. Althnian, A., et al.: Impact of dataset size on classification performance: an empirical evaluation in the medical domain. Appl. Sci. **11**(2), 796 (2021)
19. Tripathi, P., Vishwakarma, S.K., Lala, A.: sentiment analysis of English tweets using rapid miner. In: 2015 International Conference on Computational Intelligence and Communication Networks (CICN), Jabalpur, India, pp. 668–672 (2015)
20. Eisenstein, J.: Unsupervised learning for lexicon-based classification. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31, no. 1 (2017)
21. Assiri, A., Emam, A., Aldossari, H.: Arabic sentiment analysis: a survey. Int. J. Adv. Comput. Sci. Appl. **6**(12), 75–85 (2015)
22. Almujaiwel, S.: GitHub - salmujaiwel/Covid-19_1M_Saudi_Tweets: 1M Saudi Tweets on Covid-19. GitHub. https://github.com/salmujaiwel/Covid-19_1M_Saudi_Tweets. Accessed 24 July 2020
23. Alrumayyan, N., Bawazeer, S., AlJurayyad, R., Al-Razgan, M.: Analyzing user behaviors: a study of tips in foursquare. In: Alenezi, M., Qureshi, B. (eds.) 5th International Symposium on Data Mining Applications. AISC, vol. 753, pp. 153–168. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-78753-4_12
24. Github, Women Sport Arabic Tweets Dataset Dictionary (2022). https://github.com/jewelshahrani/Women-Sport-Arabic-Tweets-Dataset-Dictionary
25. Scikit learn, Support Vector Machines (2011). https://scikit-learn.org/stable/modules/svm.html
26. Alabbas, H., Al-Khateeb, M., Mansour, A., Epiphaniou, G., Frommholz, I.: Classification of colloquial Arabic tweets in real-time to detect high-risk floods. In: 2017 International Conference on Social Media, Wearable And Web Analytics, Social Media 2017, pp. 1–8 (2017)
27. Windasari, I.P., Uzzi, F.N., Satoto, K.I.: Sentiment analysis on Twitter posts: an analysis of positive or negative opinion on GoJek. In: 2017 4th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), pp. 266–269 (2017)